

Collaborative Vietnamese WordNet building using consensus quality

Trong Hai Duong¹ · Minh Quang Tran² · Thi Phuong Trang Nguyen³

Received: 7 February 2016 / Accepted: 21 July 2016 / Published online: 6 August 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Most ontologies are being developed in an engineering-oriented method: a small group of engineers carefully builds and maintains a presentation of their view of the world. Certainly, there are several tools oriented towards collaborative work: a consensus-building mechanism that allows a large group of people to contribute or annotate a common ontology in a collaborative way to reach consensus among individuals. However, the previous approaches have not yet exploited the most important problem in consensus-based collaboration, when can we get a consensus? The main goal of this research is to investigate an effective methodology for collaborative ontology building in which we apply *consensus quality* and *susceptible to consensus* to reach to the final version of the collaborative ontology building.

Keywords Collaborative ontology · Ontology · Consensus · Ontology building · Ontology engineering

1 Introduction

Human collaboration is an effort among a group of people contributing to a common goal. It can be used as the infrastructure for facilitating the creation of a common and shared understanding. Ontologies can be developed to improve the quantity and quality of communication among participants, who can then benefit from the skills and knowledge of others. Thus, it is very important and necessary for investigating and developing principle approaches and flexible tools to allow individuals to collaboratively build, refine, and integrate existing ontologies.

Most ontologies are being developed in an engineering-oriented method: a small group of engineers carefully builds and maintains a presentation of their view of the world. Maintaining such large ontologies in an engineering-oriented way is a highly complex process: developers need to regularly merge and reconcile their modifications to ensure that the ontology captures a consistent and unified view of the domain. However, conflict can lead to errors in complex ways. These errors may manifest themselves both as structural (i.e., syntactic) mismatches between developers' ontological descriptions, and as unintended logical consequences. Therefore, the tools are unsuitable for ontology construction by large groups of non-experts over the web. In other words, the previous approaches have not yet exploited the most important problem in consensus-based collaboration, which is when we can get a consensus. The main goal of this research is to investigate an effective methodology, which is using the consensus quality to not only ease the collaborative ontology building process by reducing the workload of ontology data integration but also increase the accuracy of the final version of the ontology that is based on a large number of contributors with or without domain experts.

✉ Trong Hai Duong
haiduongtrong@gmail.com

Minh Quang Tran
tranminhquang2209@gmail.com

Thi Phuong Trang Nguyen
phuongtrangict@gmail.com

¹ International University, Vietnam National University-HCMC, Ho Chi Minh City, Vietnam

² Institute of Science and Technology of Industry 4.0, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

³ Banking University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

2 Related works

According to our study, there are several tools oriented towards collaborative work [7, 8, 13, 14, 16]: a consensus-building mechanism that allows a large group of people to contribute or annotate a common ontology in a collaborative way to reach consensus among individuals. One instance of these tools is Protégé¹ which is established by Stanford University for knowledge acquisition. It provides a graphical and interactive ontology design and knowledge-based development environment. Ontology developers can access relevant information quickly, and navigate and manipulate the ontology. One of the advantages of Protégé is an open, modular design. Tudorache et al. [16] have developed Collaborative Protégé as an extension to the client–server Protégé. Collaborative Protégé allows entire groups of developers who are building an ontology collaboratively to hold discussions, chat, make annotations and make changes as a part of the ontology-development process. One of the advantages of Collaborative Protégé is the ability to create annotations. OntoWiki [2] is a web-based ontology which focuses on an instance editor that provides only rudimentary capabilities as the history of changes and ratings of ontology components. OntoWiki provides different views on instance data (e.g., a map view for geographical data or a calendar view for data containing dates). OntoEdit [15] is a collaborative ontology (CoO) editing environment that integrates numerous aspects of ontology engineering and allows multiple users to develop ontologies in three phases: a requirements specification, refinement, and evaluation/maintenance. KAON [5] focuses on changes of ontology that can cause inconsistencies, a proposed deriving evolution strategy to maintain consistencies. However, the collaborative version of aforementioned approaches may not reach to the consensus among participants since it just accepts the latest modification from any participant on collaborative process. Here, we consider a collaborative ontology building process which allows an entire group to be heard and to participate in the process of ontology building by reaching a consensus and usually aiming at completeness. The goal of collaborative process is to find a common ground and examine these issues in ontology building until mutual agreement between group members has been reached. We agree with previous works [3, 4, 7, 8, 10], there are four phases of the collaborative approach to design ontology including: (1) the preparatory phase defines the criteria, specifies boundary conditions for the ontology, and determines standards for assessing its success; (2) the anchoring phase includes the development of an initial version of the ontology which will feed the next phase (evaluation phase) while being aware and complying with the design criteria; (3) the iterative improvement phase enhances ontology until

all participants' points of view reach a consensus through a collaborative building technique. In this phase, the ontology structure will be revised and evolved by collaboration of participants. At each iterative improvement, the ontology is evaluated by aforementioned standards and conditions; (4) the application phase demonstrates the use of collaborative ontology by applying it in various ways. However, the previous approaches have not yet exploited the most important problem in consensus-based collaboration, when can we get a consensus? The main goal of this research is to investigate an effective methodology for collaborative ontology building in which we apply *consensus quality* and *susceptible to consensus* by Nguyen et al. [9] to reach to the final version of the collaborative ontology building.

3 Collaborative algorithm using consensus quality

3.1 Consensus-based collaboration overview

The Nominal Group Technique (NGT) [6] is well known as a method for decision making. It has been used to get the final result among a group whether large or small while considering all opinions and votes from group members. NGT takes into account the participants who join the discussion to choose the result. It is successful when everybody participates and understands the manners, and represents the solution or opinions by themselves without affecting the surroundings around them. NGT is a process where everyone is clearly involved and knows everything while getting the solution without missing anyone in the discussion.

One of the popular consensus-building techniques is the Delphi method [12]. This method is used for normal discussion that does not need complex communication between experts such as meeting face-to-face or having a meeting to talk at a table. It is because this method can be implemented using technology such as email or any other electronic technologies for communication where each question can be sent directly to every group expert. Even though there is a complex problem that needs to be solved, this method can be used to find the solution by sending a series of questionnaires via multiple iterations and getting a solution (data) from experts. The Delphi method is commonly used in education, to estimate forecasts and other fields. The Delphi technique can be done in four steps:

1. The moderator forms a group of experts that participates in the process to solve the problem. However, all of the experts are unidentified.
2. A person will send a questionnaire to the participants via mail or email.
3. Once the person gets the return answer from a participant, the person will analyze the results.

¹ <http://protege.stanford.edu/>.

4. At the last step, if there is no consensus reached, a combination of previous questionnaires and results will be used as a new version of the questionnaire, and the moderator will send this new version again to a participant. Step 2 is repeated until consensus is reached, or the moderator ends the process and makes a final report.

There are some different factors between these two aforementioned methods. As we already know the Delphi method is commonly used without experts needing to meet each other. In Nominal, all participants or experts need to be in one place and doing the process together. The main point of Nominal is all participants are required to meet face-to-face to reach the solution. What they believe in Nominal is, every idea or opinion is strongly agreed upon if experts or participants present their ideas formally and seriously in front of other experts. It means that in Nominal, consensus can be reached if there is real discussion. In contrast to the Delphi methods, they believe that without meeting each other and with believing the anonymous expert, the consensus result is more accurate. It is because without affecting other experts, an individual expert can find the ideas and solution based on expert knowledge, so consensus results are more reliable based on individual expertise.

3.2 Consensus quality

To solve conflicts between participants, a method following [9] has been introduced. Each participant in a collaborative group gives his or her knowledge x to a profile X , which is a set of knowledge that is collected by n participants.

$$X_h = \{x_1, x_2, \dots, x_n\}$$

where x_i is an annotation of a participant i for the object h which is the set of senses and relations of one word or phrase in Ontology-based Vietnamese WordNet.

For conflicting profiles and their consensus, a measurement has been used to evaluate these consensus which follow [9]:

$$\hat{d}(x, X) = 1 - \frac{d(x, X)}{\text{card}(X)} \quad (1)$$

where $\hat{d}(x, X)$ is the quality of consensus x in profile X . $d(x, X)$ is the sum of all different distances between an element x to the universe. $\text{card}(X)$ is the number of participants in X .

For a given distance space (U, d) , we define some parameters following [9]:

$$d_{t_mean}(X) = \frac{\sum_{x,y \in X} (d(x, y))}{k(k+1)} \quad (2)$$

$$d_x(X) = \frac{\sum_{y \in X} (d(x, y))}{k} \quad (3)$$

$$d_{\min}(X) = \min_{x \in U} d_x(X) \quad (4)$$

$$d_{\max}(X) = \max_{x \in U} d_x(X) \quad (5)$$

where $d_{t_mean}(X)$ is the total average distance of all distances in profile X . $d_x(X)$ represents the average distance of all distances between object x and the elements of profile X . $d_x(X)$ represents the average distance of all distances between object x and the elements of profile X . $d_{\min}(X)$: The minimal value of $d_x(X)$ for $x \in U$

$$\text{card}(X) = k.$$

Next, to calculate the distance between two elements of a profile X , *cosine distance* has been used as a measure. Hence, it is required to convert these elements into vectors before applying *cosine distance* function below:

$$\begin{aligned} d(x_i, x_j) &= 1 - \cos(\theta) = 1 - \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{k=1}^n A_k B_k}{\sqrt{\sum_{k=1}^n A_k^2} \sqrt{\sum_{k=1}^n B_k^2}} \end{aligned} \quad (6)$$

where $d(x_i, x_j)$ is a distance between element x_i and x_j (which $x_i, x_j \in X$), A, B are vectors of element x_i and x_j , respectively. A_k, B_k are respective components of vector A and B . $\cos(\theta)$ represents the similarity of vector A and B .

Example Let X be a profile where $X = \{2 * ab\}$; there are two votes for a and one vote for b . The above-defined values are calculated as follows:

$$d_{t_mean}(X) = \frac{2 \times (0 + 2 + 0)}{3 \times 4} = \frac{1}{3}$$

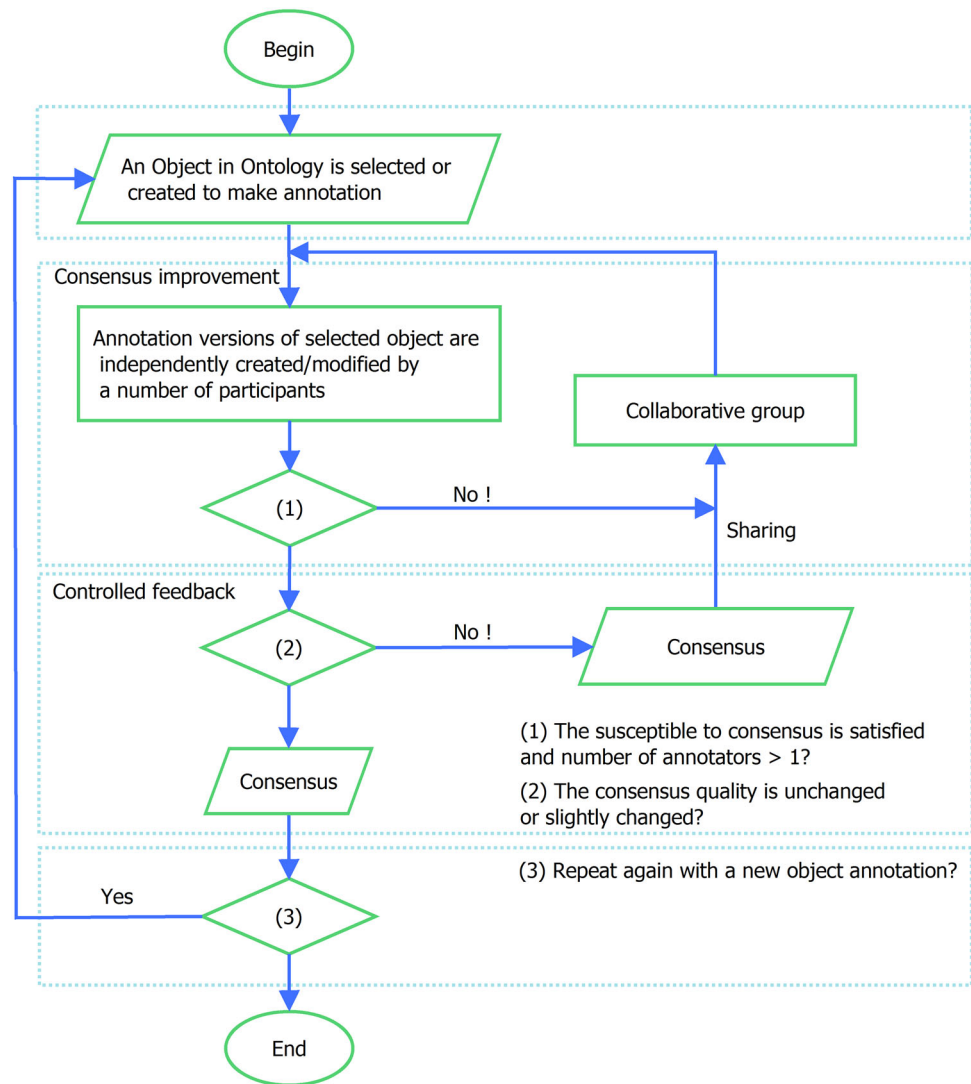
$$d_{\min}(X) = \frac{1}{3}$$

To reach an optimal profile, the inequality which follows [9] has been used. The susceptible to consensus of profile X is satisfied if and only if the following inequality takes place

$$d_{t_mean}(X) \geq d_{\min}(X) \quad (7)$$

X is susceptible to consensus (it is possible to determine a good consensus for X) if the second value is not greater than the first. Satisfying the above inequality means that the elements of profile X are dense enough for determining a good consensus. In other words, opinions represented by these elements are consistent enough for determining a good compromise.

Fig. 1 Collaborative algorithm using consensus quality



3.3 Collaborative algorithm using consensus quality

The algorithm using consensus quality is expressed in Fig. 1:

Following [4], we present features of a method for collaborative ontology building:

Phase 1 Preparatory: instead of using questionnaires in Delphi we provide criteria for ontology building [3].

Phase 2 Contribution: the changeable ontology is cloned from the original one. Participants can modify their own versions without changing the original version.

Phase 3 Consensus improvement: the annotation versions of an object are independently created/modified by a number of participants. In addition the *Susceptible to Consensus* of this object is calculated using Eqs. (2), (3) and (4). If the result satisfies the inequality (7) and the number of annotators is greater than 1, the process moves to Phase 4. Otherwise, the group keeps modifying this object until the inequality (7) occurs.

Phase 4 Controlled feedback: in this phase, the quality of consensus is calculated using Eq. (1). If consensus quality is unchanged or slightly changed, a reconciled ontology that is constructed from the integration of the generated versions will be used as a new version of the ontology. Next, this consensus version will be shared to the group and the process moves back to Phase 3. The algorithm will stop until there is no improvement that needs to be done.

Example Assume that there is a laptop which comprises four components such as CPU Memory (RAM) Hard Disk and a model name. To identify the details of this laptop we apply the algorithm as follows:

Step 1: Inviting participant to annotate the object. At the beginning there is only one participant for annotation and the result is collected as shown in Table 1.

Step 2: Calculating all possible *cosine distances* between two different points (participants).

At the moment there is only one point (participant) hence this step is not applicable.

Table 1 1-Participant annotation result

Participants	CPU	Memory	Hard Disk	Model
1	Intel i3	4 GB	500 GB	HP 4230s

Step 3: Calculating the total average distance in this profile which is $d_{t_mean}(X)$.

In (2) $d_{t_mean}(X) = 0$ (as we only have one participant)

Step 4: Calculating the minimum distance of this profile.

In (3) and (4), we have $d_{min}(X) = 0$

Step 5: Calculating the quality of consensus if the inequality of susceptible to consensus (7) occurs.

According to steps 3 and 4, the result satisfies the inequality of susceptible to consensus as $d_{t_mean}(X) = d_{min}(X) = 0$; however, the number of participants is not greater than 1.

Therefore, next, we need to increase the number of participants to 2, which let one more person to annotate the laptop, and go back to **step 1**, the result is as shown in Table 2:

Step 2: Calculating all possible *cosine distances* between two different points (participants).

To convert participants' ideas into vectors, all of the terms are counted as shown in Table 3:

As a result of Table 3, we have 2 vectors for 2 participants:

$$P1(1, 0, 1, 0, 1, 1)$$

$$P2(0, 1, 0, 1, 1, 1)$$

In (6),

$$d(P1, P2) = 1 - \frac{1 \times 0 + 0 \times 1 + 1 \times 0 + 0 \times 1 + 1 \times 1 + 1 \times 1}{\sqrt{1^2 + 0^2 + 1^2 + 0^2 + 1^2 + 1^2} \sqrt{0^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2}} = 0.5$$

Step 3: Calculating the total average distance in this profile which is $d_{t_mean}(X)$

In (2),

$$d_{t_mean}(X) = \frac{2 \times 0.5}{2 \times (2 + 1)} = 0.16$$

Step 4: Calculating the minimum distance of this profile.

In (3) and (4), we have $d_1(X) = d_2(X) = d_{min}(X) = \frac{0.5}{2} = 0.25$.

Step 5: Calculating the quality of consensus if the inequality of susceptible to consensus occurs:

According to steps 3 and 4, the result does not satisfy the inequality of susceptible to consensus (7) as $d_{t_mean}(X) < d_{min}(X)$ (due to $0.16 < 0.25$).

Therefore, we have to invite one more person to annotate this laptop to reach the consensus back to **step 1** (see Table 4):

Step 2: Calculating all possible *cosine distances* between two different points (participants).

In (6),

$$d(P1, P2) = 1 - \frac{2}{4} = 0.5$$

$$d(P1, P3) = 1 - \frac{0}{4} = 1$$

$$d(P2, P3) = 1 - \frac{1}{4} = 0.75$$

Step 3: Calculating the total average distance in this profile which is $d_{t_mean}(X)$.

In (2),

$$d_{t_mean}(X) = \frac{2 \times (0.5 + 1 + 0.75)}{3 \times (3 + 1)} = 0.375$$

Step 4: Calculating the minimum distance of this profile.

Table 2 2-Participant annotation results

Participants	CPU	Memory (GB)	Hard disk (GB)	Model
1	Intel i3	4	500	HP 4230s
2	Intel i5	8	500	HP 4230s

Table 3 Terms frequencies

Participants	Intel i3	Intel i5	4 GB	8 GB	500 GB	HP 4230s
1	1	0	1	0	1	1
2	0	1	0	1	1	1

Table 4 3-Participant annotation results

Participants	CPU	Memory (GB)	Hard disk (GB)	Model
1	Intel i3	4	500	HP 4230s
2	Intel i5	8	500	HP 4230s
3	Intel i5	6	750	HP 4530s

Table 5 4-Participant annotation results

Participants	CPU	Memory (GB)	Hard disk (GB)	Model
1	Intel i3	4	500	HP 4230s
2	Intel i5	8	500	HP 4230s
3	Intel i5	6	750	HP 4530s
4	Intel i5	8	750	HP 4530s

Table 6 5-Participant annotation results

Participants	CPU	Memory (GB)	Hard disk (GB)	Model
1	Intel i3	4	500	HP 4230s
2	Intel i5	8	500	HP 4230s
3	Intel i5	6	750	HP 4530s
4	Intel i5	8	750	HP 4530s
5	Intel i3	8	750	HP 4530s

In (3),

$$d_1(X) = 0.5$$

$$d_2(X) = 0.417$$

$$d_3(X) = 0.583$$

Then following (4), we have $d_{\min}(X) = 0.417$.

Step 5: Calculating the quality of consensus if the inequality of susceptible to consensus occurs.

According to step 3 and 4, the result does not satisfy the inequality of susceptible to consensus as $d_{t_mean}(X) < d_{\min}(X)$ (due to $0.375 < 0.417$).

Thus, we increase the number of participants to 4 and back to **step 1** again (see Table 5).

Step 2: Calculating all possible cosine distances between two different points (participants).

In (6),

$$d(P1, P2) = 0.5$$

$$d(P1, P3) = 1$$

$$d(P1, P4) = 1$$

$$d(P2, P3) = 0.75$$

$$d(P2, P4) = 0.5$$

$$d(P3, P4) = 0.25$$

Step 3: Calculating the total average distance in this profile which is $d_{t_mean}(X)$.

In (2),

$$d_{t_mean}(X) = 0.4.$$

Step 4: Calculating the minimum distance of this profile.

In (3),

$$d_1(X) = 0.625$$

$$d_2(X) = 0.4375$$

$$d_3(X) = 0.5$$

$$d_4(X) = 0.4375$$

Then following (4), we have $d_{\min}(X) = 0.4375$.

Step 5: Calculating the quality of consensus if the inequality of susceptible to consensus occurs.

To be consistent with steps 3 and 4, the result does not satisfy the inequality of susceptible to consensus (7) as $d_{t_mean}(X) < d_{\min}(X)$ (due to $0.4 < 0.4375$).

As a result, we invite one more participant to annotate this laptop and back to **step 1** (see Table 6).

Step 2: Calculating all possible distances between two different points (participants).

In (6),

$$d(P1, P2) = 0.5$$

$$d(P1, P3) = 1$$

$$d(P1, P4) = 1$$

$$d(P1, P5) = 0.75$$

$$d(P2, P3) = 0.75$$

$$d(P2, P4) = 0.5$$

$$d(P2, P5) = 0.75$$

$$d(P3, P4) = 0.25$$

$$d(P3, P5) = 0.5$$

$$d(P4, P5) = 0.25$$

Step 3: Calculating the total average distance in this profile which is $d_{t_mean}(X)$.

In (2),

$$d_{t_mean}(X) = 0.417$$

Step 4: Calculating the minimum distance of this profile. In (3),

Table 7 The relation in Vietnamese WordNet

Property	Domain	Range	Target
hyponymOf	Synset	Synset	Nouns, Adjs
Entails	Synset	Synset	Verbs
similarTo	Synset	Synset	Adjectives
memberMeronymOf	Synset	Synset	Nouns
substanceMeronymOf	Synset	Synset	Nouns
partMeronymOf	Synset	Synset	Nouns
classifiedByTopic	Synset	Synset	Nouns, Adjs, Verbs
classifiedByUsage	Synset	Synset	Nouns, Adjs, Verbs
classifiedByRegion	Synset	Synset	Nouns, Adjectives, Verbs
causes	Synset	Synset	Verbs
sameVerbGroupAs	Synset	Synset	Verbs
attribute	Synset	Synset	Nouns to Adjectives
derivationallyRelated	WordSense	WordSense	Nouns, Verbs, Adjectives, Adverbs
antonymOf	WordSense	WordSense	Nouns, Verbs, Adjectives, Adverbs
seeAlso	WordSense	WordSense	Verbs, Adjectives
participleOf	WordSense	WordSense	Adjectives to Verbs
adjectivePertainsTo	Synset	Synset	Adjectives to Nouns or Adjectives
adverbPertainsTo	Synset	Synset	Adverbs to Adjectives
gloss	WordSense	xsd: string	Synset and Sentence
frame	Verb-WordSense	xsd: string	Synset and a verb construction pattern
partOf	Synset	Synset	Nouns
originalSenseOf	Synset	Synset	Nouns, Verbs
vietEng	Synset	Synset	Nouns, Verbs, Adverbs, Adjectives

$$d_1(X) = 0.65$$

$$d_2(X) = 0.5$$

$$d_3(X) = 0.5$$

$$d_4(X) = 0.4$$

$$d_5(X) = 0.5625$$

Then following (4) we have $d_{\min}(X) = 0.4$.

Step 5: Calculating the quality of consensus if the inequality of susceptible to consensus occurs.

To be compatible with steps 3 and 4, the result satisfies the inequality of susceptible to consensus (7) as $d_{t_mean}(X) > d_{\min}(X)$ (due to $0.417 > 0.4$).

Finally, this consensus is shared to everyone who has annotated the laptop. The quality of consensus of the first round is computed as below:

In (1),

$$\begin{aligned}
 \hat{d}(x, X) &= 1 - \frac{d(x, X)}{\text{card}(x)} \\
 &= 1 - \frac{0.65 + 0.5 + 0.5 + 0.4 + 0.5625}{5} \\
 &= 0.4775.
 \end{aligned}$$

4 Experiment

4.1 Vietnamese WordNet

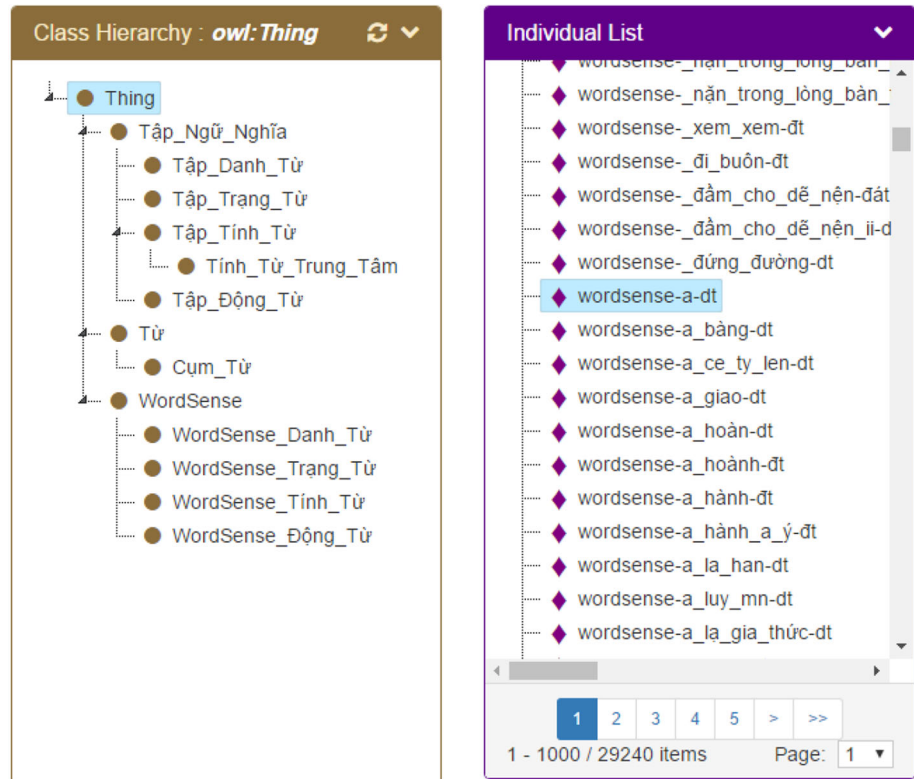
Our proposed approach is assessed by applying for collaborative Vietnamese WordNet building. The structure and relations of Vietnamese WordNet (VW) are initially derived from the English WordNet [1]. VW classifies most of words in Vietnamese language into four main types including Noun–Verb–Adjective and Adverb. These words are put into different type of synsets which stands for synonym sets and interconnected by a number of various relationships. Regarding the structure, VW has three main classes consisting of *Synset*, *Word* and *WordSense*. *Synset* and *WordSense* have subclasses based on the distinction of lexical groups. *Synset* has four subclasses containing *NounSynset*, *VerbSynset*, *AdjectiveSynset*, and *AdverbSynset*. *WordSense* has four subclasses including *NounWordSense*, *VerbWordSense*, *AdjectiveWordSense*, and *AdverbWordSense*. *Word* has a subclass *Collocation* which is used to store words or phrases in Vietnamese. The class hierarchy of VW is inherited from WordNet [4] and the properties and its significance are shown in Table 7.

Fig. 2 Sample XML dictionary data

```

<?xml version="1.0"encoding="UTF-8"
standalone="yes"?>
<dictionary>
  <word>
    <name>a</name>
    <type>dt.</type>
    <definition>Từ thứ nhất trong bản chữ cái.
    một nguyên âm: A hoa. a thường.</definition>
  </word>
</dictionary>

```

Fig. 3 The Vietnamese WordNet's class hierarchy

4.2 Demonstration

Creating Vietnamese WordNet Ontology.

To initialize the first version of VW, there are three steps that need to be done as follows:

Step 1—Extracting raw data and converting it to semi-structure data.

We used Từ Điển Tiếng Việt, which is according to Vietnamese–Vietnamese dictionary, and extracted all of the words inside the dictionary to an XML formatted file. Basically, there are three details of a word which are extracted such as name, type and definition. The format of this XML looks like the following (see Fig. 2):

Step 2—Cleansing the extracted data.

A cleansing process is performed before adding all of the words in XML file to the ontology as it is not always certain that the extracted data are correct. To be more detailed, sometimes a word name could not be retrieved accurately and a blank or a symbol is returned instead. Therefore, these incorrect words are removed or ignored.

Step 3—Matching words with ontology classes and adding them to VW.

This step is to define which types of words match with the classes in Vietnamese Ontology. In the initializing version, the VW is built in a simple way where we only select 4 types of word to be added up to its respective classes: ‘Danh từ’ (Noun) will be individualized in ‘WordSense_Danh_Từ’, ‘Động từ’ (Verb) will be individualized in ‘WordSense_Động_Từ’, ‘Trạng từ’ (Adverb) will

Fig. 4 Original version of a word

Individual description for *wordsense-an_tâm-tt*

Original Version Collaborative Version Your Version Other Users Versions

Display Name
wordsense-an_tâm-tt

IRI
http://www.semanticweb.org/phuongtrang/ontologies/2015/8/untitled-ontology-56#wor

Annotations 1

gloss [language: vi] ✕
Yên lòng. không lo ngại: Nói cho ông ấy an tâm.

Description

Types 1
WordSense_Tinh_Từ ✕

Same Individual As 0

Different Individuals 0

Property Assertions

Object Property Assertions 0

Data Property Assertions 0

Negative Object Property Assertions 0

Fig. 5 User version of a word

Individual description for *wordsense-an_tâm-tt*

Clone Other Version

Annotations + 1

gloss [language: vi] ✕
Yên lòng. không lo ngại: Nói cho ông ấy an tâm.

Description

Types + 1
WordSense_Tinh_Từ ✕

Same Individual As + 0

Different Individuals + 0

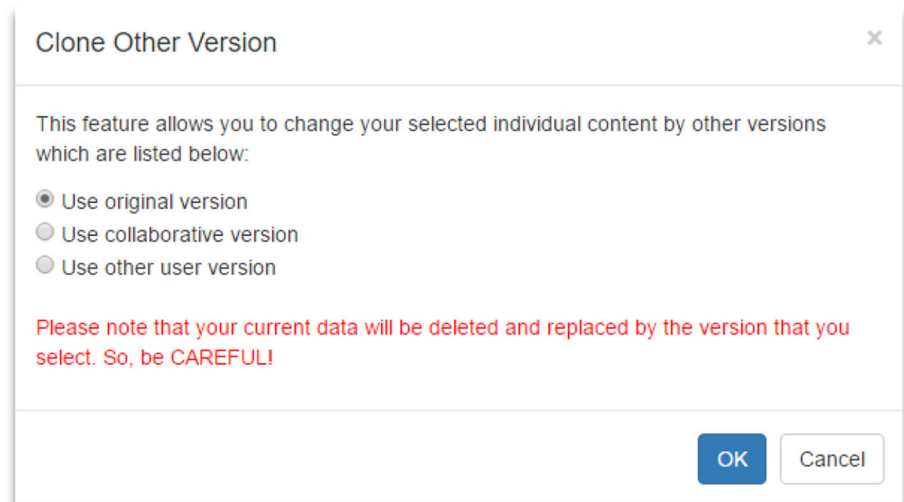
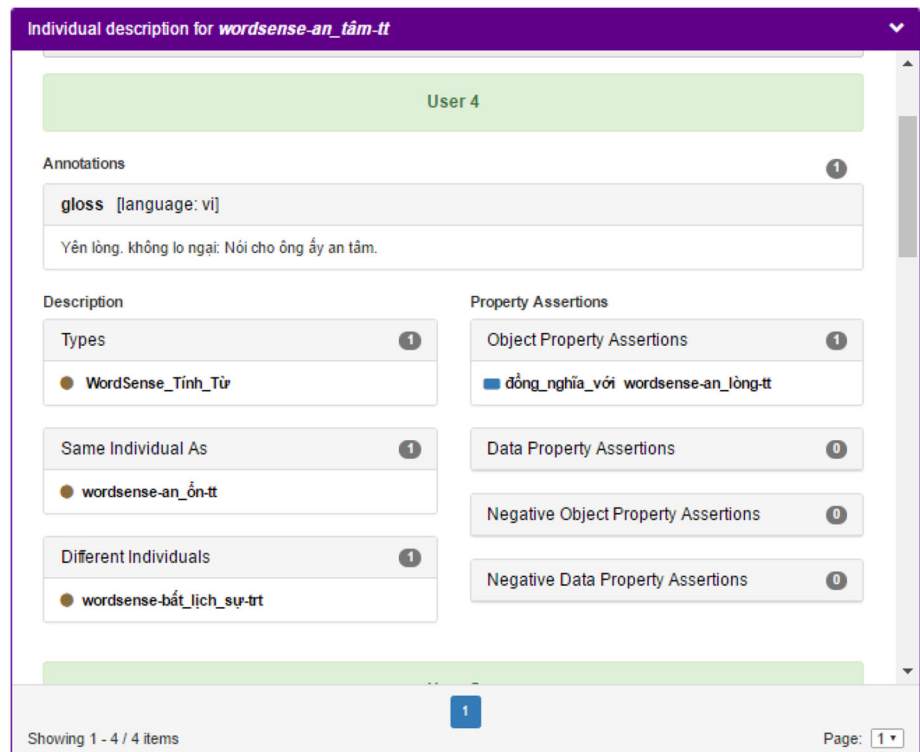
Property Assertions

Object Property Assertions 0
+

Data Property Assertions + 0

Negative Object Property Assertions + 0

Negative Data Property Assertions + 0

Fig. 6 ‘Clone’ feature options**Fig. 7** An example of ‘Other User Versions’ feature/tab (1)

be individualized in ‘*WordSense_Trạng_Từ*’, and ‘*Tính từ*’ (Adjective) will be individualized in ‘*WordSense_Tính_Từ*’. Other types of word will be individualized in ‘*OWL:Thing*’, which includes cảm từ, số từ, giới từ, đại từ, thán từ. After that the XML file is parsed and the result is added to the Vietnamese WordNet.

The final result—the first Vietnamese Ontology

Finally, the Vietnamese Wordnet Ontology is initialized with 29240 individuals, which ‘*WordSense_Danh_Từ*’ class

has 12679 individuals, ‘*WordSense_Trạng_Từ*’ class has 2863 individuals, ‘*WordSense_Tính_Từ*’ class has 4030 individuals, and ‘*WordSense_Động_Từ*’ class has 0 individuals. Other individuals are not classified and by default, they are individuals of ‘*OWL:Thing*’ class. Fig. 3 illustrates the initialized version of VW.

However, this version of VW only contains words along with their definitions and there is no connection/or relation-

Fig. 8 An example of ‘Other User Versions’ feature/tab (2)

The screenshot displays the 'Individual description for wordsense-an_tâm-tt' interface. At the top, a green bar identifies 'User 2'. Below this, the 'Annotations' section shows a 'gloss' with the value '[language: vi]' and a Vietnamese sentence: 'Yên lòng. không lo ngại: Nói cho ông ấy an tâm.' The 'Description' section includes 'Types' (with 'WordSense_Tinh_Từ' selected) and 'Same Individual As' and 'Different Individuals' (both with counts of 0). The 'Property Assertions' section lists 'Object Property Assertions' (with 'đồng_nghĩa_với wordsense-an_lòng-tt' selected), 'Data Property Assertions' (0), 'Negative Object Property Assertions' (0), and 'Negative Data Property Assertions' (0). A green bar at the bottom identifies 'User 3'. The footer shows 'Showing 1 - 4 / 4 items' and 'Page: 1'.

ship between words. As a result, to completely build the ontology, lots of collaborative work need to be done.

Collaborative Vietnamese WordNet Building

To improve this VW, we build a web-based application, called Ontology Wiki (OntWiki), to upload and display all details of VW, by which users are able to view the class hierarchy, object properties hierarchy, data properties hierarchy, annotation properties hierarchy and individual listed by class. In addition, the OntWiki also allows users to view multiple versions of an individual. There are four types of version such as ‘Original Version’ which shows the original version of an individual in the OWL file. ‘User Version’ is the user’s opinions of an individual, user can use this feature to submit their point of view. ‘Others Users Versions’ are the versions of multiple users who have given their ideas on the same individual, and finally, ‘Collaborative Version’ automatically integrates all versions of all users to create a collaborative version, which makes use of our proposed collaborative algorithm using consensus quality. In this experiment, we select 500 individuals of ‘WordSense_Tinh_Từ’ and share them with thirty participants (collaborative group).

First of all, an administrator of OntWiki created thirty accounts and gave to this collaborative group. The administrator is the only one who is able to modify the structure of VW, as well as the original version of individuals (see Fig. 4). Normal users can only perform personal idea submissions of individuals, which they can only modify their versions, but not original version and other users’ versions.

Next, to use OntWiki, participants need to login, then select VW, which is already uploaded by an administrator, select ‘WordSense_Tinh_Từ’ class in individual page, and choose a provided list of individuals and start working on it. To ease up the initialization stage of user version, OntWiki provides a ‘clone’ feature that allows users to reuse or copy the original version, collaborative version, or other user version to their own (see Figs. 5, 6). There are eight characteristics of an individual that user can modify in their own version, which includes ‘Annotations’, ‘Types’, ‘Same Individual As’, ‘Different Individuals’, ‘Object Property Assertions’, ‘Data Property Assertions’, ‘Negative Object Property Assertions’ and ‘Negative Data Property Assertions’.

After a period of time, the final results will be collected and administrators will start to upgrade all of shared individuals using the collaborative versions. By this way, the results are always transparent between users (see Figs. 7, 8); therefore, the effectiveness goes up very much, time consuming is reduced due to no meeting conduction, and the effort given is not high and also has high-quality output.

5 Conclusion

In this work, an effective methodology for collaborative ontology building is improved from [3, 7, 8] using quality of consensus [9] to reach consensus among participants in col-

laborative group. A susceptible to consensus is to answer that when we can have a consensus and the quality of consensus is to determine if the final version of the collaborative ontology has been reached or not. We applied the proposed method for Vietnamese WordNet building. In future work, we will combine trust-based consensus [4] and quality of consensus to solve leading problem in collaboration.

Acknowledgements This research is funded by International University, Vietnam National University, Ho Chi Minh City under grant number T2016-01-IT/HD-DHQT-QLKH.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Assem, M.V., Gangemi, A., Schreiber, G.: RDF/OWL representation of WordNet. <http://www.w3.org/TR/wordnet-rdf/>
2. Auer, S., Dietzold, S., Riechert, T.: OntoWikia tool for social semantic collaboration. In: The 5th International Semantic Web Conference, pp. 736–749. Springer, ISWC Athens GA (2006)
3. Duong, T.H., Jo, G.S.: Collaborative ontology building by reaching consensus among participants. *Int. J. Inf.* **13**(5), 1557–1569 (2010)
4. Duong, T.H., Nguyen, N.T., Nguyen, D.C., Nguyen, T.P.T., Selamat, A.: Trust-based consensus for collaborative ontology building. *Cybern. Syst.* **45**(2), 146–164 (2014)
5. Gabel, T., Sure, Y., Voelker, J.: KAON – ontology management infrastructure. D3.1.1.a, SEKT Project: Semantically Enabled Knowledge Technologies (2004)
6. Gallagher, M., Hares, T., Spencer, J., Bradshaw, C., Webb, I.: The nominal group technique: a research tool for general practice? *Fam. Pract.* **10**, 76–81 (1993)
7. Holsapple, C.W., Joshi, K.D.: Collaborative approach in ontology design. *Commun. ACM* **45**, 42–47 (2002)
8. Karapiperis, S.: Consensus building in collaborative ontology engineering processes. *J. Univ. Knowl. Manag.* **1**, 199–216 (2006)
9. Nguyen, N.T.: Advanced Methods for Inconsistent Knowledge Management. Springer, London (2008)
10. Nguyen, Q.U., Duong, T.H., Kang, S.: Solving conflict on collaborative knowledge via social networking using consensus choice. In: Proceedings of the 4th International Conference on Computational Collective Intelligence: Technologies and Applications, Vol. Part I, pp. 21–30 (2012)
11. Noy, N., McGuinness, D.L.: Ontology development a guide to creating your first ontology. KSL Technical Report KSL-01-05 (2001)
12. Pill, J.: The Delphi method: substance context a critique and an annotated bibliography. *Soc. Econ. Plan. Sci.* **5**, 57–71 (1971)
13. Ruiz, E.J., Grau, B.C., Horrocks, I., Berlanga, R.: Building ontologies collaboratively using contentCVS. In: Proceedings of the 22nd International Workshop on Description Logics (2009)
14. Ruiz, E.J., Grau, B.C., Horrocks, I., Berlanga, R.: Conflict detection and resolution in collaborative ontology development technical report (2009) Tool and user study available at <http://www.kronoactujies/people/Ernst/contentcvs>
15. Sure Y.: OntoEdit: collaborative ontology development for the semantic web. In: International Semantic Web Conference 2002 (ISWC 2002) Sardinia Italy, pp. 221–235 (2002)
16. Tudorache, T., Noy, N.F., Tu, S.W., Musen, M.A.: Supporting collaborative ontology development in Protg. In: Proceedings of 7th International Semantic Web Conference Karlsruhe, pp. 17–32. Springer, Germany (2008)